枫香叶片变色期全长转录组测序及分析

刘雄盛，尹国平，肖玉菲，蒋燚，王仁杰，黄荣林，姜英，王勇*

（广西壮族自治区林业科学研究院，广西优良用材林资源培育重点实验室，南宁 530002）

摘要：枫香因其树形优美，入秋后叶色红艳或橙黄，极具观赏价值，是优良的景观生态树种。为了解枫香叶片变色及其次级代谢过程的遗传基础，该文以枫香 5 个变色期叶片混合样品为材料，利用单分子实时测序技术（PacBio 平台）对其进行全长转录组测序。结果表明：（1）全长转录组测序共获得 41.04 Gb 的高质量数据，从中鉴定出全长非嵌合序列 563 180 条，通过聚类和去冗余，获得 27 269 条高质量全长转录本。在 27 269 条全长转录本中预测到 2 035 条长链非编码 RNA（lncRNA），并检测出 14 892 个简单重复序列（SSR）位点和 1 856 个转录因子。（2）基因注释结果表明，NR、GO、COG、KEGG 等 8 个数据库共注释了 24 857 条转录本，KEGG 数据库共获得了 124 个条代谢途径，主要有核糖体、碳代谢、氨基酸生物合成等，在类黄酮和叶绿素代谢途径中分别有 49 和 71 个转录本参与。上述结果初步揭示了枫香叶片变色期转录组信息以及功能特性，为后续研究枫香叶片变色分子机制、色素代谢合成途径和调控、相关功能基因克隆以及叶色改良提供基础数据。

关键词：枫香，叶片变色期，单分子实时测序技术，全长转录组，基因功能注释

中图分类号：Q781；S718.43　　　　**文献标识码：**A

Sequencing and analysis of full-length transcriptome from *Liquidambar formosana* leaves in discoloration stage

LIU Xiongsheng, YIN Guoping, XIAO Yufei, JIANG Yi, WANG Renjie, HUANG Ronglin, JIANG Ying, WANG Yong*

(*Guangxi Key Laboratory of Superior Trees Resource Cultivation*, *Guangxi Zhuang Autonomous Region Forestry Research Institute*, Nanning 530002, China)

Abstract: *Liquidambar formosana* is an excellent landscape ecological tree species because its beautiful tree shape and red or orange leaves in autumn. In order to understand the genetic basis of discoloration and secondary metabolism of *L. formosana* leaves, the mixed samples of *L. formosana* leaves at five discoloration periods were used for full-length transcriptome sequencing using single-molecule real-time sequencing technique (PacBio platform). The results were as follows: (1) High-quality 41.04 Gb data were obtained by full-length transcriptome sequencing, from which 563 180 full-length non-chimeric sequences were identified, and 27 269 high-quality full-length transcripts were obtained by clustering and de-redundancy. In 27 269 full-length transcripts, 2 035 long-chain non-coding RNA (lncRNA) were predicted, and 14 892 simple repeat sequence (SSR) sites and 1 856 transcription factors were detected. (2) The results of gene annotation showed that a total of 24 857 transcripts were annotated in 8 databases such as NR, GO, COG and KEGG, and 124 metabolic pathways were obtained in KEGG database, including ribosome, carbon metabolism, amino acid biosynthesis and so on. 49 and 71 transcripts were involved in flavonoid and chlorophyll metabolism respectively. The above results preliminarily reveal the transcriptome information and functional characteristics of *L. formosana* leaves during

基金项目：广西优良用材林资源培育重点实验室自主课题（2019-A-03-02）；广西林业科技推广示范项目（桂林科研[2021]26 号）[Supported by Independent Project of Guangxi Key Laboratory for Cultivation of Excellent Timber Forest Resources (2019-A-03-02); Guangxi Forestry Science and Technology Extension Demonstration Project (No. 26 [2021] Guilin Scientific Research)]。

第一作者：刘雄盛（1988-），硕士，副研究员，主要从事林木遗传育种研究，（E-mail）517261654@qq.com。

***通信作者：**王勇，硕士，高级工程师，主要从事森林生态研究，（E-mail）12084474@qq.com。

the discoloration period, and provide basic data for the follow-up study of the molecular mechanism of discoloration, the pathway and regulation of pigment metabolism and synthesis, the cloning of related functional genes and the improvement of leaf color.

Key words: *Liquidambar formosana*, leaf discoloration stage, single-molecule real-time sequencing technology, full-length transcriptome, gene function annotation

随着人们对观赏植物需求的日益增加，彩叶树种因其色相丰富、色泽艳丽、观赏价值高等特点备受关注（王振兴等，2016）。彩叶树种叶色变化机制及其影响因素也随之成为了研究的热点（李卫星等，2017）。近年来，学者们对彩叶树种叶色变化做了大量研究，研究内容逐渐从叶片变色过程中表型形态和细胞结构变化（杜文文等，2019；梁玲等，2020）、生态适应意义（Menzies et al., 2016；陈颖卓和黄至欢，2016）和生理生化特征（Junker & Ensminger, 2016；冯露等，2017；赵东辉等，2019）到叶色变化分子调控机制（陆小雨等，2020；Gao et al., 2021）。叶片内叶绿素、花青素、类胡萝卜素等色素含量变化是彩叶植物叶色变化的直接原因（Jiang et al., 2016；李卫星等，2017）。目前，叶色变化分子机制研究主要集中在叶绿素、类胡萝卜素和花青素等色素合成与调控有关基因、转录因子、miRNA（MicroRNA）及其靶基因的表达模式和功能等方面（Guan et al., 2014；Yang et al., 2015；Li et al., 2015；Gao et al., 2020），为植物叶色形成的遗传和基因组研究、叶色调控以及彩叶植物资源开发奠定了基础。

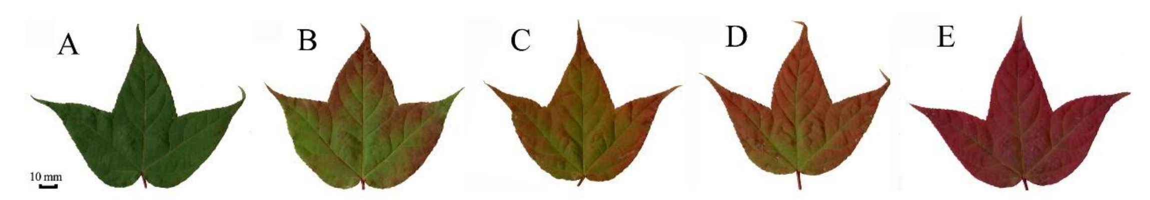
枫香（*Liquidambar formosana*）隶属金缕梅科（Hamamelidaceae）枫香树属（*Liquidambar*），为落叶乔木，主要分布于中国秦岭和淮河以南各省区海拔 1 000 m 以下的低山次生林内，在越南北部、老挝和朝鲜南部亦有分布（黄宁等，2021）。枫香适应性强，天然易更新，适宜红、黄壤土，有荒山先锋树种之称（罗紫东等，2016）。入秋后，枫香叶片逐渐变为红色或橙黄，极具观赏价值（王冬雪等，2017），然而，在不同区域枫香叶色表现差异较大，具明显区域性，推测枫香叶色可能具有特异的调控机制。目前，对于枫香叶色变化的研究主要集中叶色变化过程中光合作用（罗紫东等，2016）、色素含量（刘儒等，2017；王冬雪等，2019）等生理生化方面。在分子生物学方面，Wen 等（2014）利用二代测序技术对枫香秋叶衰老时基因表达谱进行了研究，并对秋季变色和叶片衰老过程中可能的基因调控进行了讨论。然而，由于二代测序技术读取序列短，拼接时无法提供长转录本，且会丢失可变剪接等重要信息。因此，目前，对于枫香叶片变色分子机制的研究仍然缺乏遗传信息，限制了枫香叶色资源的开发利用。

PacBio 单分子实时测序（single-molecule real-time, SMRT）技术测序读长远超 Illumina 等二代测序技术，因此可以对完整的 mRNA 直接进行从头测序，从而得到转录本的全长信息，具有发现更多可变剪切序列和新功能基因，改善基因组注释，鉴定更多的 lncRNA（long non-coding RNA）以及准确定位融合基因等特点（Tian et al., 2018；夏丽飞等，2020），广泛应用于转录组学、表观遗传学和大型基因组组装等领域（赵陆滟等，2019；吴志铭等，2020）。夏丽飞等（2020）利用 PacBio 单分子实时测序技术获得紫鹃茶树（*Camellia sinensis* var. *asssamica*）全长转录本信息，为其变色机制研究提供基础数据。Jia 等（2020）对高山杜鹃（*Rhododendron lapponicum*）进行全长转录组测序，获得 75 002 个高质量全长转录本，为其花色形成机制研究提供参考。本实验通过 PacBio 的单分子实时测序技术进行枫香叶片变色期全长转录组测序，获取高质量全长转录本进行分析和功能注释，旨在为后续研究枫香叶片变色分子机制、色素代谢合成途径和调控、相关功能基因克隆以及叶色改良提供遗传基础。

1 材料与方法

1.1 实验材料

枫香叶片变色期材料采自广西壮族自治区百色市德保县红叶森林公园（23°21′19″ E、106°39′5″ N），采用平均木法，选取 5 株生长健壮的枫香植株。在每株枫香树上选取东、南、西、北 4 个方向的枝条做好标记，自 2018 年 9 月底开始，每 15~20 d 采集样品 1 次，每次采样时在每株枫香的每个枝条上采集 5 片完整叶片进行混合，共采集了 5 次样品。各时期叶片颜色如图 1 所示。



A. 第 1 次采样；**B**. 第 2 次采样；**C**. 第 3 次采样；**D**. 第 4 次采样；**E**. 第 5 次采样。

A. The first sampling; **B**. The second sampling; **C**. The third sampling; **D**. The fourth sampling; **E**. The fifth sampling.

图 1 不同变色期枫香叶片

Fig. 1 Leaf of *Liquidambar formosana* at different discoloration stages

1.2 RNA 提取和 cDNA 文库构建

采用 Trizol 试剂提取枫香各时期叶片样品的总 RNA（ribonucleic acid），用 1.2%的琼脂糖凝胶电泳检测 RNA 降解和污染情况。使用 NanoDrop 2000 分光光度计测量 RNA 的纯度、浓度和吸收峰。进一步用 Aligent Bioanalyzer 2100 检测 RNA 质量。RNA 检测合格后，取各时期 28S/18S＞1，且 RIN＞6.5 的 RNA 等量混合，使用 SMARTer™ PCR cDNA Synthesis Kit 合成 mRNA（messenger RNA）的全长 cDNA（complementary DNA），通过 PCR（polymerase chain reaction）扩增放大全长 cDNA 进行末端修复，并连接 SMRT（single molecule real-time）哑铃型接头进行核酸外切酶消化，获得一个 1~6 kb 的文库。在 Pacific Bioscience RS II 平台上进行 SMRT 测序（委托百迈客生物科技有限公司）。

1.3 测序数据的质量控制和序列聚类

将原始序列中长度＜50 bp 的片段和准确性＜0.90 的序列过滤，获取到过滤后的测序数据。根据序列中的接头，将序列转换成环形一致序列 CCS（circular con-sensus），再根据 CCS 判断是否有 3'引物、5'引物以及 PolyA，将序列分成全长和非全长序列。将来自同一转录本的全长序列聚类，相似的聚成一簇，每个簇得到一条一致序列，校正后，获得用于后续分析的高质量序列（丁玉梅等，2020）。

1.4 转录组完整性评估和结构分析

使用 CD-HIT（Cluster Database at High Identity with Tolerance）软件去除转录本中的冗余序列，获得非冗余转录本序列；利用 BUSCO（Benchmarking Universal Single-Copy Orthologs）对去冗余后的转录本进行完整性评估；使用 TransDecoder 软件进行转录本编码区序列及对应氨基酸序列预测；应用 CPC（Coding Potential Calculator）分析、CNCI（Coding-Non-Coding Index）分析、Pfam（Protein Families）蛋白结构域分析、CPAT（Coding Potential Assessment Tool）分析 4 种方法预测 lncRNA；筛选 500 bp 以上的转录本，利用 MISA（MIcroSAtellite identification tool）软件做 SSR（Simple Sequence Repeat）分析。

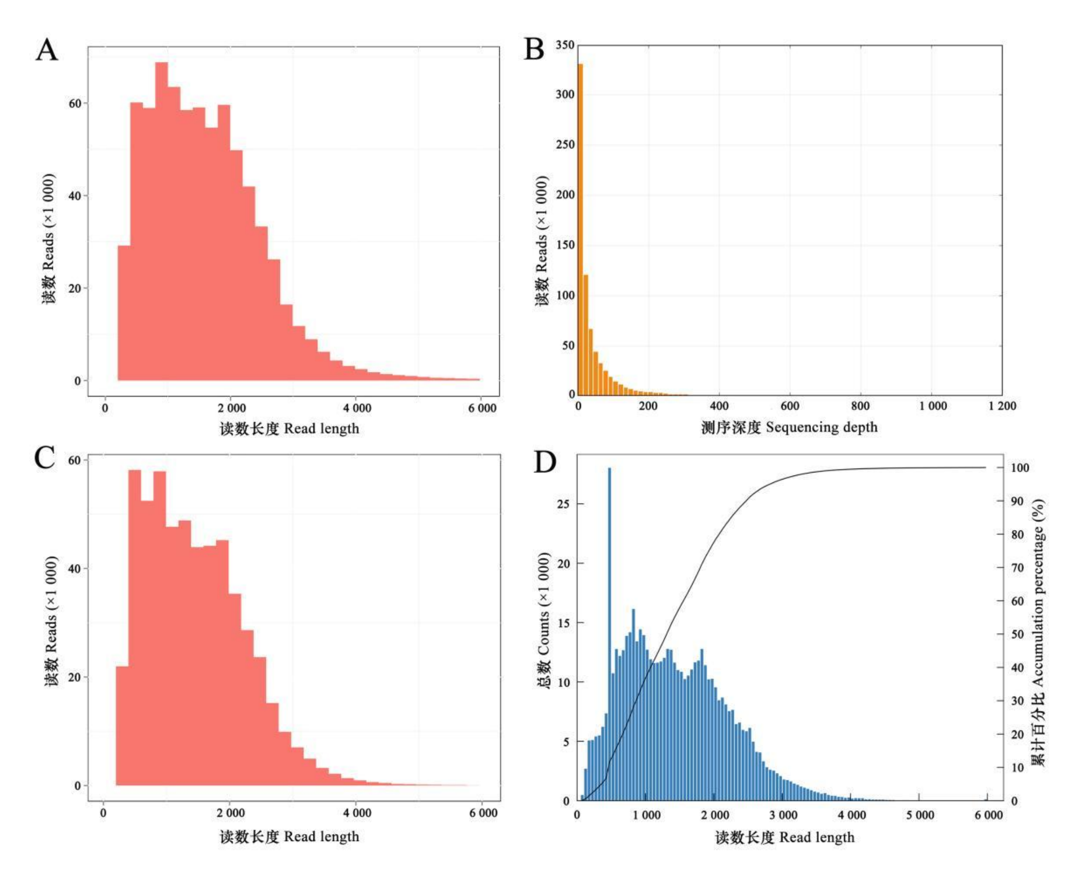
1.5 转录本功能注释

使用 BLAST（Basic Local Alignment Search Tool）软件将得到的非冗余转录本序列与 NR（Non-Redundant Protein Sequence Database）、Swissprot（Swiss-Prot Protein Sequence Database）、GO（Gene Ontology）、COG（Clusters of Orthologous Groups of proteins）、KOG（Clusters of Orthologous Groups for Eukaryotic Complete Genomes）、eggNOG（Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups Database）、Pfam、KEGG（Kyoto Encyclopedia of Genes and Genomes）数据库进行比对，获得转录本的注释信息（邹智等，2021）；利用 iTAK（Plant Transcription factor & Protein Kinase Identifier and Classifier）软件鉴定转录因子。

2 结果与分析

2.1 枫香叶片变色期全长转录组文库构建

使用 2 个 SMRT cell 进行全长转录组测序，获得 41.04 Gb 清洁数据。从中提取到 731 849 条 CCS，约计 1 178 416 098 bp。CCS 平均长度为 1 610 bp（图 2：A），测序平均深度为 46×（图 2：B）。从 731 849 条 CCS 中鉴定出全长非嵌合序列 FLNC（full length readsnon-chimeric）563 180 条，占比为 76.95%，其长度分布如图 2：C 所示。对 FLNC 进行聚类，获得 50 736 条一致序列，校正后，共得到 50 282 条（99.11%）高质量一致序列，长度分布如图 2：D 所示。对获得的高质量一致序列去冗余，得到 27 269 条全长转录本。



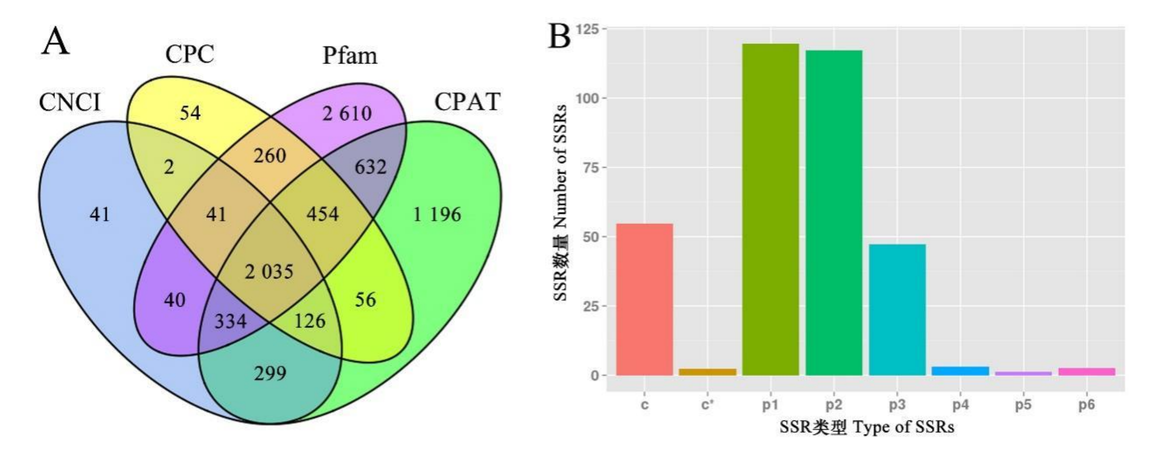
A. CCS 长度分布；**B**. 生成 CCS 的 full passes 数分布；**C**. FLNC 长度分布；**D**. 高质量一致序列长度分布。
A. Length distribution of CCS; **B**. Distribution of full passes for CCS extraction; **C**. Length distribution of FLNC; **D**. Length distribution of polished high-quality consensus isoforms.

图 2 枫香叶片变色期全长转录组序列分布

Fig. 2 Sequence distribution of the full-length transcriptome in leaf discoloration stage of *Liquidambar formosana*

2.2 编码区序列、LncRNA 和 SSR 分析

通过编码区序列及对应氨基酸序列的预测，共获得 25 408 个开放阅读框 ORF（open reading frame），其中 20 281 条 ORF 是完整的，占比 79. 57%。lncRNA 预测分析中，CPC、CNCI、CPAT 和 Pfam 分别预测到 3 028、2 848、5 132 和 6 406 条 lncRNA，2 035 条为共有序列（图 3：A）。利用 MISA 软件筛选 500 bp 以上的转录本，共获得 14 892 个 SSR 位点，其中单碱基 SSR 有 5 124 个，数目最多，其平均密度约为每 Mb43.5 个（图 3：B）。



A. lncRNA 预测维恩图；**B**. SSR 密度分布；**c**. 混合 SSR；**c***. 带有重合碱基的混合 SSR；**p1**. 单碱基 SSR；**p2**. 双碱基 SSR；**p3**. 三碱基 SSR；**p4**. 四碱基 SSR；**p5**. 五碱基 SSR；**p6**. 六碱基 SSR。

A. Venn diagram of predicted lncRNA; **B**. Density distribution of SSR; **c**. Compound SSR; **c***. Compound SSR with overlapping bases; **p1**. Mono-nucleotide SSR; **p2**. Di-nucleotide SSR; **p3**. Tri-nucleotide SSR; **p4**. Tetra-nucleotide SSR; **p5**. Pentanucleotide SSR; **p6**. Hexa-nucleotide SSR.

图 3 枫香叶片变色期转录本的 lncRNA 预测和 SSR 分析

Fig. 3 lncRNA prediction and SSR analysis of transcripts in leaf discoloration stage of *Liquidambar formosana*

2.3 转录本功能注释

8 个数据库的功能注释结果显示（表 1），共注释 24 857 条序列，占总转录本的 91.15 %。在 NR 注释的物种中，葡萄（*Vitis vinifera*）占比最高，为 36.39 %，其次是可可树（*Theobroma cacao*，7.09 %）、荷花（*Nelumbo nucifera*，6.71 %）、麻枫树（*Jatropha curcas*，3.41 %）等（图 4）。

表 1 注释转录本的统计情况

Table 1 Statistics of transcripts annotated

所注释数据库 Annotated database	转录本数量 Number of transcripts
NR	24 726
GO	17 535
COG	10 809
KEGG	10 666
KOG	15 286
Pfam	20 455
Swiss-Prot	17 910
eggNOG	24 158
合计 Total	24 857

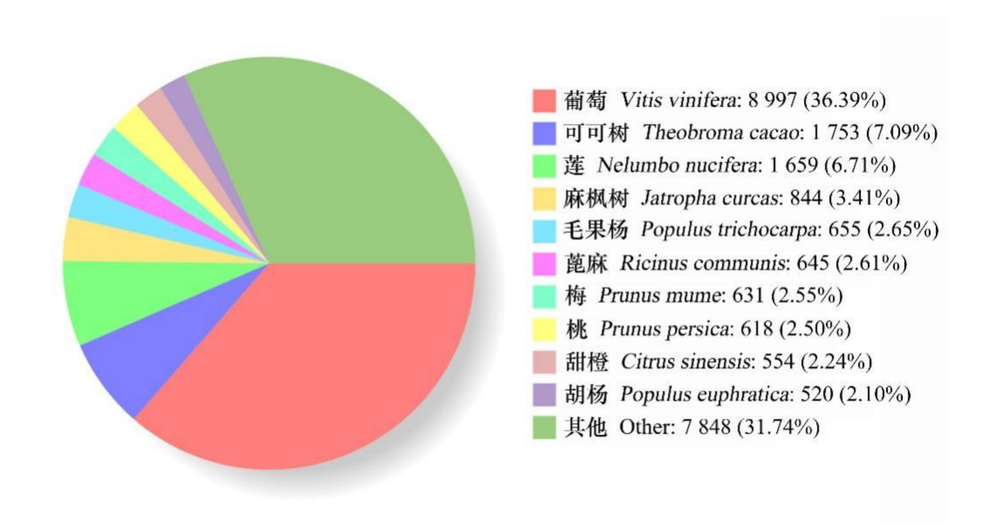
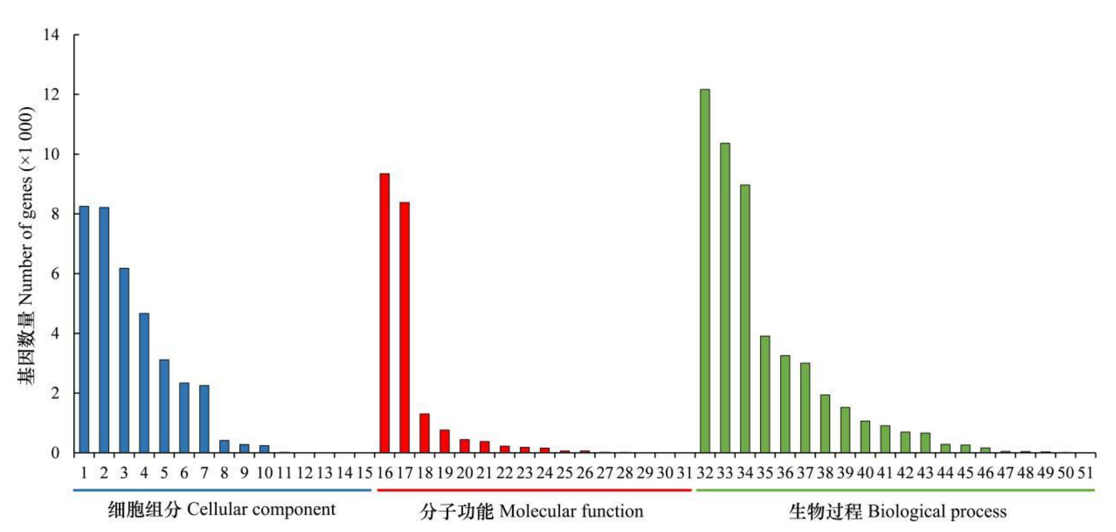


图 4 NR 注释同源种分布统计

Fig. 4 Statistics of NR homologous species distribution

GO 注释的 17 535 个转录本中，包含生物过程（biological process，49 314）、细胞组分（cellular component，36 008）、分子功能（molecular function，21 366）3 大类 51 个亚类。其中，生物过程中代谢过程（metabolic process）和细胞过程（cellular process）占比较高，分别为 69.4%和 59.1%；细胞组分中细胞区域（cell part）和细胞（cell）占比较高，分别为 47.1%和 46.8%；分子功能中催化活性（catalytic activity）和结合功能（binding）占比较高，分别为 53.3%和 47.8%（图 5）。



1. 细胞组分；**2**. 细胞；**3**. 细胞器；**4**. 膜结构；**5**. 细胞器部分；**6**. 高分子配合物；**7**. 细胞膜组分；**8**. 胞外区域部分；**9**. 细胞连接；**10**. 膜封闭内腔；**11**. 类核；**12**. 胞外区部分；**13**. 病毒体；**14**. 病毒体组分；**15**. 细胞外基质；**16**. 催化活性；**17**. 结合活性；**18**. 转运活性；**19**. 结构分子活性；**20**. 电子载体活性；**21**. 核酸结合转录因子活性；**22**. 分子转导活性；**23**. 抗氧化活性；**24**. 酶调节活性；**25**. 受体活性；**26**. 结合蛋白转录活性；**27**. 蛋白标签；**28**. 营养贮存活性；**29**. 金属伴侣蛋白活性；**30**. 核酸交换因子活性；**31**. 翻译调节活性；**32**. 代谢进程；**33**. 细胞进程；**34**. 单组织生物过程；**35**. 应激反应；**36**. 生物调节；**37**. 定位；**38**. 细胞组织部分；**39**. 发育进程；**40**. 多细胞进程；**41**. 信号传导；**42**. 有机体进程；**43**. 繁殖进程；**44**. 免疫系统进程；**45**. 生长；**46**. 繁殖；**47**. 节律进程；**48**. 生物相；**49**. 生物附着；**50**. 细胞活动；**51**. 细胞失活。

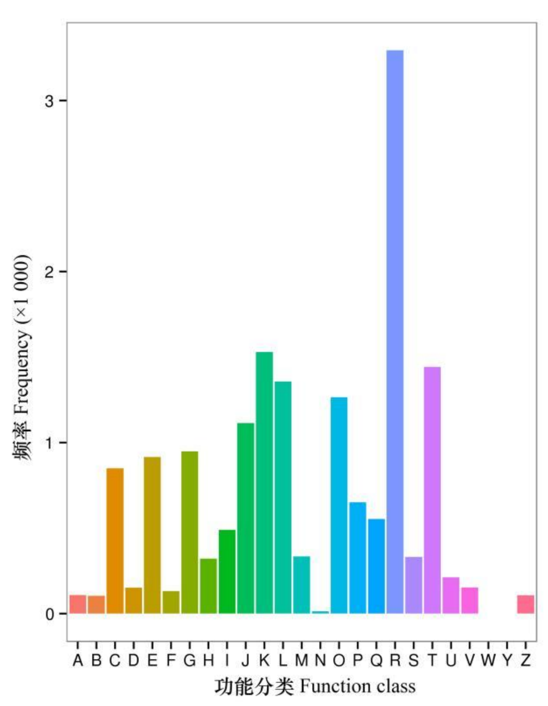
1. Cell part; **2**. Cell; **3**. Organelle; **4**. Membrane; **5**. Organelle part; **6**. Macromolecular complex; **7**. Membrane part;

8. Extracellular region; **9**. Cell junction; **10**. Membrane-enclosed lumen; **11**. Nucleoid; **12**. Extracellular region part; **13**. Virion; **14**. Virion part; **15**. Extracellular matrix; **16**. Catalytic activity; **17**. Binding; **18**. Transporter activity; **19**. Structural molecule activity; **20**. Electron carrier activity; **21**. Nucleic acid binding transcription factor activity; **22**. Molecular transducer activity; **23**. Antioxidant activity; **24**. Enzyme regulator activity; **25**. Receptor activity; **26**. Protein binding transcription factor activity; **27**. Protein tag; **28**. Nutrient reservoir activity; **29**. Guanyl-nucleotide exchange factor activity; **30**. Metallochaperone activity; **31**. Translation regulator activity; **32**. Metabolic process; **33**. Cellular process; **34**. Single-organism process; **35**. Response to stimulus; **36**. Biological regulation; **37**. Localization; **38**. Cellular component organization or biogenesis; **39**. Developmental process; **40**. Multicellular organismal process; **41**. Signaling; **42**. Multi-organism process; **43**. Reproductive process; **44**. Immune system process; **45**. Growth; **46**. Reproduction; **47**. Rhythmic process; **48**. Biological phase; **49**. Biological adhesion; **50**. Locomotion; **51**. Cell killing.

图 5 转录本 GO 功能分类统计

Fig. 5 Transcript GO function classification statistics

COG 注释中，一般功能预测（general function prediction only，20.11%）所占比例最高，其次为转录（transcription，9.34%）、转录信号转导机制（signal transduction mechanisms，8.81 %）、复制、重组和生物发生（replication, recombination and biogenesis，8.29%）等（图6）。



A. RNA 加工与修饰：108（0.66%）；**B**. 染色体结构和活力：104（0.64%）；**C**. 能量生成和转换：849（5.18%）；**D**. 细胞周期控制、蛋白质折叠和分子伴侣：152（0.93%）；**E**. 氨基酸运输和代谢：915（5.59%）；**F**. 核苷酸运输和代谢：131（0.80%）；**G**. 碳水化合物运输和代谢：949（5.80%）；**H**. 辅酶运输和代谢：321（1.96%）；**I**. 脂质运输和代谢：490（2.99%）；**J**. 翻译、核糖体结构和生物合成：1 114（6.80%）；**K**. 转录：1 530（9.34%）；**L**. 重复、重组和修饰：1 357（8.29%）；**M**. 细胞壁膜生物合成：334（2.04%）；**N**. 细胞运动：14（0.09%）；**O**. 翻译后修饰、蛋白质折叠和分子伴侣：1 265（7.72%）；**P**. 矿脂运输和代谢：651（3.98%）；**Q**. 次生代谢物合成、运输和代谢：553（3.38%）；**R**. 一般功能预测：3 293（20.11%）；**S**. 功能未知：332（2.03%）；**T**. 信号传导：1 442（8.81%）；**U**. 细胞内转运、分泌和小泡运输：212（1.29%）；**V**. 防卫机制：153（0.93%）；**W**. 胞外结构：0（0.00%）；**Y**. 核结构：0（0.00%）；**Z**. 细胞构架：107（0.65%）。

A. RNA processing and modification: 108 (0.66%); **B**. Chromatin structure and dynamics: 104 (0.64%); **C**. Energy

production and conversion: 849 (5.18%); **D**. Cell cycle control, cell division, chromosome partitioning: 152 (0.93%); **E**. Amio acid transport and metabolism: 915 (5.59%); **F**. Nucleotide transport and metabolism: 131 (0.80%); **G**. Carbohydrate transport and metabolism: 949 (5.80%); **H**. Coenzyme transport and metabolism: 321 (1.96%); **I**. Lipid transport and metabolism: 490 (2.99%); **J**. Translation, ribosomal structure and biogenesis: 1 114 (6.80%); **K**. Transcription: 1 530 (9.34%); **L**. Replication, recombination and repair: 1 357 (8.29%); **M**. Cell wall/membrane/envelope biogenesis: 334 (2.04%); **N**. Cell motility: 14 (0.09%); **O**. Posttranslational modification, protein turnover, chaperones: 1 265 (7.72%); **P**. Inorganic ion transport and metabolism: 651 (3.98%); **Q**. Secondary metabolites biosynthesis, transport and catabolism: 553 (3.38%); **R**. General function prediction only: 3 293 (20.11%); **S**. Function unknown: 332 (2.03%); **T**. Signal transduction mechanisms: 1 442 (8.81%); **U**. Intracellular trafficking, secretion, and vesicular transport: 212 (1.29%); **V**. Defense mechanisms: 153 (0.93%); **W**. Extracellular structures: 0 (0.00%); **Y**. Nuclear structure: 0 (0.00%); **Z**. Cytoskeleton: 107 (0.65%).

图 6 COG 注释分类

Fig. 6 Classification of COG annotation

KEGG 数据库中共注释 10 666 个转录本，根据代谢途径分为机体系统（organismal systems，289）、代谢（metabolism，6 596）、遗传信息处理（genetic information processing，2 920）、环境信息处理（environmental information processing，304）、细胞进程（cellular processes，542）等 5 个一级代谢通路和 18 个二级代谢通路以及 124 个三级代谢通路（图 7）。其中碳水化合物代谢（carbohydrate metabolism，1 919）、翻译（translation，1 735）、全局和概述地图（global and overview maps，1 318）、氨基酸代谢（amino acid metabolism，1 167）、折叠，分类，降解（folding, sorting and degradation，1 044）等二级代谢通路注释的转录本较多。三级代谢通路中核糖体（ribosome，693）、碳代谢（carbon metabolism，617）、氨基酸的生物合成（biosynthesis of amino acids，473）、内质网中的蛋白质加工（protein processing in endoplasmic reticulum，437）、剪接体（spliceosome，348）、糖酵解/糖异生（glycolysis/gluconeogenesis，296）、氧化磷酸化（oxidative phosphorylation，292）、RNA 运转（RNA transport，269）、光合生物体中的碳固定（carbon fixation in photosynthetic organisms，239）、植物激素信号转导（plant hormone signal transduction，235）等注释的转录本较多。

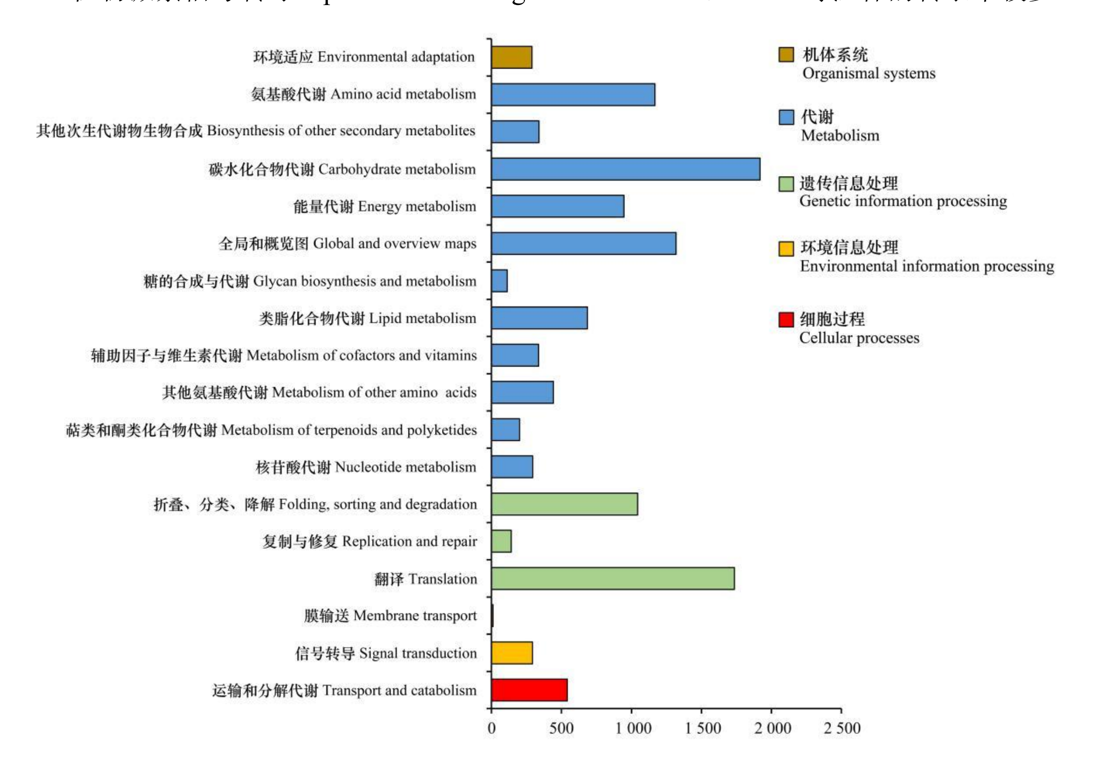


图 7 KEGG 注释分类

Fig. 7 Classification of KEGG annotation

2.4 转录因子

使用 iTAK 软件共预测得到转录因子 1 856 个，分属 159 个基因家族。其中，RLK-Pelle_LRK10L-2 家族转录因子最多，为 97 个，其次为 RLK-Pelle_DLSV，为 66 个。图 8 显示了转录因子数排名前 20 的基因家族。

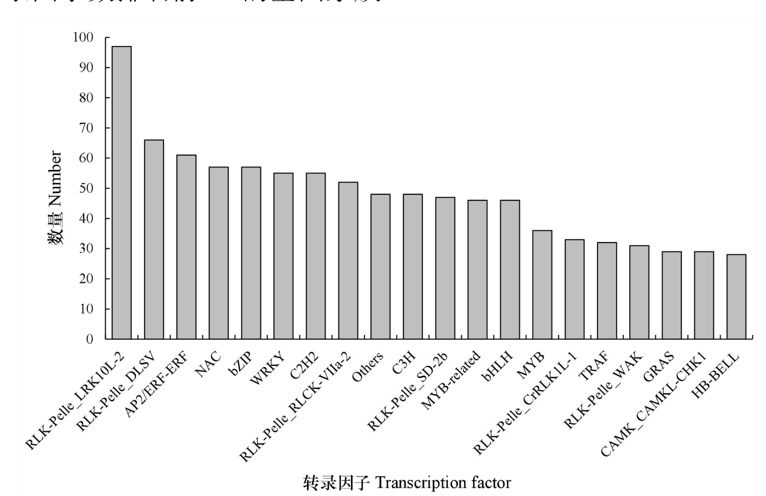


图 8 不同转录因子家族成员分布情况

Fig. 8 Distribution of different transcription factor family members

3 讨论与结论

近年来，SMRT 测序已成为全长转录组研究的最可靠、有效的策略，特别是对于没有参考基因组序列的非模式植物（潘敏等，2020；邹智等，2021）。本研究应用 SMRT 测序技术，在 PacBio RS II 平台上对枫香叶片变色期进行了研究，总共产生了 41.04 Gb 的测序数据，从中共提取到 731 849 条全长序列，全长非嵌合序列占 76.95%，测序质量较好，能够满足后续挖掘基因信息的需要。SMRT 测序技术获得的转录本的长度比下一代高通量测序技术获得的转录本的长度要长(Jia et al., 2020)。本研究中，枫香叶片变色期转录本的平均长度为 1 610 bp，远高于 Wen 等（2014）利用 Illumina 测序技术得到的枫香绿叶和红叶转录本长度（165 bp)，这说明 PacBio SMRT 测序技术是获取转录本序列，特别是获取长转录本序列的有效方法。

LncRNA 是一类转录本长度大于 200 nt 的 RNA 分子，参与调节植物的发育和生长、次生代谢和植物的逆境反应（Liu et al., 2019）。本研究中，我们用 4 种方法在枫香变色期全长转录组中获得 2 035 条 lncRNAs，这些 lncRNAs 将为进一步研究枫香叶色变化分子机制奠定基础。SMRT 测序获得的转录本开发 SSR 标记是一种有效可行的方法（夏丽飞等，2020）。本研究分析枫香叶片变色期全长转录组，共检测到 14 892 个 SSR 位点，单碱基 SSR 数量最多。枫香分布范围广，抗逆性强，叶片呈色特异。因此，上述 SSR 位点为为枫香遗传多样性研究、比较基因组学研究、基因作图研究、种群遗传学研究和其他类型的遗传研究提供有价值的遗传工具（李文燕等，2020；Wu et al., 2020）。

在 8 个数据库中，有 24 857 个枫香叶片变色期转录本通过序列比对进行了注释，注释转录本比例为 91.15 %，远高于 Wend 等(2014)利用二代测序技术注释的转录本比例(56%)。这表明本研究鉴定到大量枫香叶片中的基因。剩下的 2 412 个转录本没有 BLAST 匹配，可能代表了枫香叶片特异的基因或未知基因。Nr 注释结果表明，枫香叶片全长转录组序列信息与葡萄最相似（36.39%），与 Wend 等（2014）的研究结果一致。GO、COG 和 KEGG 分

类结果表明，大量转录本参与转录、复制、重组和修复，并具有催化活性，具有不同的分子功能，有 10 666 份转录本被分配到特定的途径，参与多种生物学途径。因此，我们的研究结果为进一步开展枫香叶色变化的分子研究提供了丰富的遗传信息。

植物叶片内叶绿素、花青素以及类胡萝卜素含量比例和分布决定了叶片的颜色，而色素代谢主要受结构基因和转录因子调控（Becker et al., 2014；李卫星等，2017；陈璇等，2020）。相关研究表明，叶绿素含量降低，花色素苷大量积累是导致枫香叶片变红的直接原因（刘儒等，2017；Yin at al., 2022）。前人研究表明，*HEMA*1（Glutamyl-tRNA reductase 1）、*CAO*（Chlorophyllide a oxygenase）等基因是调控叶绿素合成的重要基因（Wu et al., 2007），*NYC*1（Non yellow coloring 1）和 *NOL*（Nonyellow coloring1-like）基因在叶绿素降解过程中起关键作用（Sato et al., 2007），HD-Zip、WRKY 和 GATA 家族的转录因子可调控叶绿素含量（An et al., 2014; 李卫星等，2017）；在本研究的叶绿素代谢途径中，有 27 个转录本被注释为 *HEMA*（Glutamyl-tRNA reductase）、*PPOX*（Protoporphyrinogen oxidase）、*CHLD*（Magnesium chelatase subunit D）、*CHLM*（Magnesium chelatase subunit M）、*POR*（Light-dependent protochlorophyllide reductase）、*CAO*、*NYC*1、*NOL*、*HCAR*[7-Hydroxymethylchlorophylla (hmchl) reductase]基因，有 17、55、19 个转录本分别属于 HD-Zip、WRKY 和 GATA 家族。尤其是在叶绿素降解过程中起关键作用的 *NYC*1 和 *NOL* 基因，可能是调控枫香叶片叶绿素含量的关键基因。

C4H（Cinnamate 4-hydroxylase）、CHS（Chalcone synthase）、F3H（Flavanone 3-hydroxylase）、F3′H（Flavonoid 3'-hydroxylase）、F3′5′H（Flavonoid-3',5'-hydroxylase）、DFR（Dihydroflavonol-4-reductase）和 ANS（Anthocyanidin synthase）是调控植物花青素生物合成的关键酶，直接影响花青素合成（许倩等, 2020；Jia et al., 2020），MYB 和 bHLH 家族的转录因子通过调节花青素生物合成中基因的表达起着关键作用（刘恺媛等，2021）。本研究中，有 49 个转录本参与类黄酮生物合成途径，在花青素合成途径中，有 31 个转录本被注释为编码 C4H、HCT（Shikimic acid/Quinic acid hydroxy cinnamyl transferase）、CYP98A（Cytochrome P450）、C3′H（p-coumaroyl shikimate/quinate 3′-hydroxylas）、CHS、F3H、CYP75B1（Cytochrome P450 75B1）、DFR、ANS 的关键基因，有 36 和 46 个转录本分别属于 MYB 和 bHLH 家族。这些参与叶绿素代谢和花青素生物合成的基因将有助于后续进一步理解枫香的叶色调控机制。

综上所述，本研究枫香叶片变色期全长转录组测序共获得 41.04 Gb 的高质量数据，获得 27 269 条高质量全长转录本，并注释了 24 857 条转录本，还预测到 2 035 条 lncRNA，检测出 14 892 个 SSR 位点和 1 856 个转录因子。初步揭示枫香叶片变色期转录组信息以及功能特性，为后续开展枫香叶色变化分子调控机制以及叶色改良研究提供基础数据。

参考文献：

AN Y, HAN X, TANG S, et al., 2014. Poplar GATA transcription factor PdGNC is capable of regulating chloroplast ultrastructure, photosynthesis, and vegetative growth in *Arabidopsis* under varying nitrogen levels[J]. Plant Cell, 119(2): 313-327.

BECKER C, KLAERING HP, KROH LW, et al., 2014. Cool-cultivated red leaf lettuce accumulates cyanidin-3-O-(6-O-malonyl)-glucoside and caffeoylmalic acid[J]. Food Chem, 146: 404-411.

CHEN YZ, HUANG ZH, 2016. A mini review on adaption of young leaf redness[J]. Biodivers Sci, 24(9): 1062-1067. [陈颖卓，黄至欢，2016. 红色幼叶的适应意义探讨[J]. 生物多样性，24(9): 1062-1067.]

CHEN X, XIE J, YUE YZ, et al., 2020. Advances in research on leaf coloration mechanism of colored leaf plants[J]. Acta Bot Boreal–Occident Sin, 40(2): 358-364. [陈璇，谢军，岳远征，等，2020. 彩叶植物叶片呈色分子机制研究进[J]. 西北植物学报，40(2): 358-364.]

DU WW, CUI GF, WANG JH, et al., 2019. Effects of foliar variegation structure on leaf color in

Begonia gulinqingensis[J]. Guihaia, 39(6): 812-820. [杜文文，崔光芬，王继华，等，2019. 古林箐秋海棠叶斑结构对叶色的影响[J]. 广西植物，39(6): 812-820.]
DING YM, GAO T, BAO HH, et al., 2020. The identifcation and key genes analysis of NBS type disease-resistance gene from *Cucurbita fcifotia*[J]. Plant Physiol J, 56(9): 1833-1844. [丁玉梅，高婷，暴会会，等，2020. 黑籽南瓜 NBS 类抗病基因的鉴别及关键基因分析[J]. 植物生理学报，56(9): 1833-1844.]
FENG L, WU JY, JU YQ, et al., 2017. Leaf color characteristics and photosynthetic characteristics of purple-leafed *Lagerstroemia indica* 'Ebony Ember'[J]. J Beijing For Univ, 39(12): 93-101. [冯露，吴际洋，鞠易倩，等，2017. 紫叶紫薇呈色生理及光合特性研究[J]. 北京林业大学学报，39(12): 93-101.]
GAO TM, WEI SL, CHEN J, et al., 2020. Cytological, genetic, and proteomic analysis of a sesame (*Sesamum indicum* L.) mutantSiyl-1 with yellow-green leaf color[J]. Genes Genom, 42(1): 25-39.
GAO YF, ZHAO DH, ZHANG JQ, et al., 2021. De novo transcriptome sequencing and anthocyanin metabolite analysis reveals leaf color of *Acer pseudosieboldianum* in autumn[J]. BMC Genomics, 22(1): 383.
GUAN XY, PANG MX, NAH G, et al., 2014. MiR828 and miR858 regulate homoeologous MYB2 gene functions in *Arabidopsis* trichome and cotton fibre development[J]. Nat Commun, 5(1): 3050-3059.
HUANG N, LIU GN, YANG JS, et al., 2021. Effects of gibberellin concentration and soaking seed time on seed germination of *Liquidambar formosana* Hance[J]. Seed, 40(3): 97-101. [黄宁，刘革宁，杨继生，等，2021. 赤霉素浓度和浸种时间对枫香种子萌发的影响[J]. 种子，40(3): 97-101.]
JIA XP, TANG L, MEI XY, et al., 2020. Single-molecule long-read sequencing of the full-length transcriptome of *Rhododendron lapponicum* L.[J]. Sci Rep, 10(1): 6755.
JIANG XR, PENG JG, JIA MX, et al., 2016. Relationship between leaf reddening, ROS and antioxidants in *Buxus microphylla* during overwintering[J]. Acta Physiol Plant, 38: 199.
JUNKER LV, ENSMINGER I, 2016. Relationship between leaf optical properties, chlorophyll fluorescence and pigment changes in senescing *Acer saccharum* leaves[J]. Tree Physiol, 36(6): 694-711.
LI WY, CUI BY, LIU QJ, 2020. Full-length transcriptome sequencing, annotation and SSR prediction of Job's tears (*Coix lachryma-jobi*) seedling leaves[J]. Mol Plant Breed, 18(15): 4855-4870. [李文燕，崔百元，刘勤坚，2020. 薏苡苗期叶片全长转录组测序、注释及 SSR 预测[J]. 分子植物育种，18(15): 4855-4870.]
LI WX, YANG XB, HE ZC, et al., 2017. Research advances in the regulatory mechanisms of leaf coloration[J]. Acta Hortic Sin, 44(9): 1811-1824. [李卫星，杨舜博，何智冲，等，2017. 植物叶色变化机制研究进展[J]. 园艺学报，44(9): 1811-1824.]
LI Y, ZHANG ZY, WANG P, et al., 2015. Comprehensive transcriptome analysis discovers novel candidate genes related to leaf color in a *Lagerstroemia indica* yellow leaf mutant[J]. Genes Genom, 37(10): 851-863.
LIANG L, HUANG YQ, CHEN XH, 2020. Anatomical structure and pigment content of *Davidia involucrate* leaves and bracts with different colors[J]. Acta Bot Boreal-Occident Sin, 40(9): 1539-1548. [梁玲，黄玉琼，陈小红，2020. 不同色彩珙桐叶片和苞片解剖结构及色素含量比较研究[J]. 西北植物学报，40(9): 1539-1548.]
LIU KY, WANG ML, XIN HB, et al., 2021. Anthocyanin biosynthesis and regulate mechanisms in plants: a review[J]. Chin Agric Sci Bull, 37(14): 41-51. [刘恺媛，王茂良，辛海波，等，2021. 植物花青素合成与调控研究进展[J]. 中国农学通报，37(14): 41-51.]

LIU R, YUAN QQ, YUAN XP, et al., 2017. The relationship with change of pigment content in leaves of different *Liquidambar formosana* families and change of leaf color[J]. S Chin For Sci, 45(4): 46-49. [刘儒，原勤勤，袁小平，等，2017. 不同枫香家系叶片色素含量变化及其与叶色变化的关系[J]. 南方林业科学，45(4): 46-49.]

LIU S, WU L, QI HR, et al., 2019. LncRNA/circRNA–miRNA–mRNA networks regulate the development of root and shoot meristems of *Populus*[J]. Ind Crop Prod, 133: 333-347.

LU XY, CHEN Z, TANG F, et al., 2020. Combined transcriptomic and metabolomic analysis reveals mechanism of anthocyanin changes in red maple (*Acer rubrum*) leaves[J]. Sci Silv Sin, 56(1): 38-54. [陆小雨，陈竹，唐菲，等，2020. 转录组与代谢组联合解析红花槭叶片中花青素苷变化机制[J]. 林业科学，56(1): 38-54.]

LUO ZD, GUAN HD, ZHANG XP, et al., 2016. Changes in photosynthetic capacity during leaf senescence of *Liquidambar formosana*[J]. Chin J Appl Ecol, 27(10): 3129-3136. [罗紫东，关华德，章新平，等，2016. 枫香叶片衰老过程中光合能力的变化[J]. 应用生态学报，27(10): 3129-3136.]

MENZIES IJ, YOUARD LW, LORD JM, et al., 2016. Leaf colour polymorphisms: a balance between plant defence and photosynthesis[J]. J Ecol, 104(1): 104-113.

PAN M, YU XD, CAI ZP, et al., 2020. Transcriptome data analysis of *Artocarpus heterophyllus* stems andleaves[J]. Chin J Trop Crop, 41(7): 1288-1297. [潘敏，于旭东，蔡泽坪，等，2020. 菠萝蜜茎叶全长转录组分析[J]. 热带作物学报，41(7): 1288-1297.]

SATO Y, MORITA R, NISHIMURA M, et al., 2007. Mendel's gene cotyledon gene encodes a positive regulator of the chlorophyll-degrading pathway[J]. Proc Natl Acad Sci USA, 104(35): 14169-14174.

TIAN JY, FENG SJ, LIU YL, et al., 2018. Single-molecule long-read sequencing of *Zanthoxylum bungeanum* Maxim. transcriptome: identification of aroma-related genes [J]. Forests, 9(12): 765-777.

WANG DX, DE YJ, SHI X, et al., 2017. Change of leaf color of four *Liquidambar formosana* families under different environmental conditions[J]. For Res, 30(3): 503-510. [王冬雪，德永军，施翔，等，2017. 4 个枫香家系在不同环境下的叶色变化[J]. 林业科学研究，30(3): 503-510.]

WANG DX, SUN HJ, DE YJ, et al., 2019. Change of leaf color of *Liquidambar formosana* seedlings under different light quality treatments[J]. For Res, 32(4): 158-164. [王冬雪，孙海菁，德永军，等，2019. 不同光质处理对枫香幼苗叶色的影响[J]. 林业科学研究，32(4): 158-164.]

WANG ZX, YU YF, CHEN L, et al., 2016. Advances in leaf pigment composition, structure and photosynthetic characteristics of colored-leaf plants[J]. Plant Physiol J, 52(1): 1-7. [王振兴，于云飞，陈丽，等，2016. 彩叶植物叶片色素组成、结构以及光合特性的研究进展[J]. 植物生理学报，52(1): 1-7.]

WEN CH, LIN SS, CHU FH, 2014. Transcriptome analysis of a subtropical deciduous tree: autumn leaf senescence gene expression profile of Formosan Gum[J]. Plant Cell Physiol, 56(1): 163-174.

WU ZM, ZHANG X, HE B, 2007. A chlorophyll deficient rice mutant with impaired chlorophyllide esterification in chlorophyll biosynthesis[J]. Plant Physiol, 145(1): 29-40.

WU QC, ZANG FQ, XIE XM, et al., 2020. Full-length transcriptome sequencing analysis and development of EST-SSR markers for the endangered species *Populus wulianensis*[J]. Sci Rep, 10(1): 16249.

WU ZM, FENG YH, YANG ZQ, 2020. Preliminary analysis of SNPs related to rosin formation in *Pinus massoniana* based on transcriptome sequencing[J]. Guangxi For Sci, 49(2): 157-160. [吴志铭，冯源恒，杨章旗，2020. 基于马尾松转录组测序的产脂相关 SNPs 初步分析[J]. 广西林业科学，49(2): 157-160.]

XIA LF, SUN YN, SONG WX, et al., 2020. Full-length transcriptome analysis of Zijuan tea (*Camellia sinensis* var. *asssamica* (Masters) Kitamura) base on pacBio platform[J]. Genom Appl Biol, 39(6): 2646-2658. [夏丽飞，孙云南，宋维希，等，2020. 基于 PacBio 平台的

紫娟茶树全长转录组分析[J]. 基因组学与应用生物学，39(6): 2646-2658.]
XU Q, ZHANG C, WU JW, et al., 2020. Research progress in biosynthesis of anthocyanins[J]. Chem Ind For Prod, 40(3): 1-11. [许倩，张晨，吴嘉维，等，2020. 花青素的生物合成研究进展[J]. 林产化学与工业，40(3): 1-11.]
YANG YX, CHEN XX, XU B, et al., 2015. Phenotype and transcriptome analysis reveals chloroplast development and pigment biosynthesis together influenced the leaf color formation in mutants of *Anthurium andraeanum* 'Sonate'[J]. Front Plant Sci, 6: 139-155.
ZHAO DH, GAO YF, RONG LP, et al., 2019. Study on physiological change of *Acer pseudo-sieboldianum* leaf color in autumn[J]. Non-wood For Res, 37(2): 114-119. [赵东辉，高玉福，荣立苹，等，2019. 紫花槭秋季叶片呈色生理变化研究[J]. 经济林研究，37(2): 114-119.]
ZHAO LY, CAO SY, LONG YS, et al., 2019. Applications and research progresses of full-length transcriptome sequencing in plants[J]. J Plant Genet Resour, 20(6): 1390-1398. [赵陆滟，曹绍玉，龙云树，等，2019.全长转录组测序在植物中的应用研究进展[J]. 植物遗传资源学报，20(6): 1390-1398.]
ZOU Z, ZHAO YG, ZHANG L, et al., 2021. Single-molecule real-time (SMRT)-based full-length transcriptome analysis of tigernut (*Cyperus esculentus* L.)[J]. Chin J Oil Crop Sci, 43(2): 229-235. [邹智，赵永国，张丽，等，2021. 基于单分子实时测序的油莎豆全长转录组分析[J]. 中国油料作物学报，43(2): 229-235.]